

A Submission to the Phoenix Challenge 2005 Award Paper Program

The CENTAUR System: Helping to Protect the NIPRNet

Presenters

- Marc I. Kellner, Ph.D <mik@cert.org> – CERT® Network Situational Awareness team, Software Engineering Institute, Carnegie Mellon University
- Jeffrey J. Jaime, Capt., USAF <jeffrey.jaime@tic.dod.mil> – Applied Technology Unit, Joint Task Force - Global Network Operations, United States Strategic Command

Executive Summary

This paper describes the capabilities of the CENTAUR system, which has been developed to help DoD information operations security analysts better understand and defend the NIPRNet. CENTAUR is the largest system for Global Situational Awareness of the NIPRNet available to Tier 1 Computer Network Defense Analysts. It has been deployed and used routinely over the past 2+ years by dozens of analysts at JTF-GNO/NetDefense (formerly DoD-CERT), NSA, and most recently at Service CERTs (e.g., AFNOSC) and regional CERTs (e.g., CONUS). The CENTAUR system maintains a repository of detailed data regarding network traffic handled by the border and backbone routers on the NIPRNet, as far as April 2002.

The CENTAUR system provides users with powerful and flexible capabilities to perform exploration and analysis of this NIPRNet traffic data – both current and historical. CENTAUR is not yet another system for automatically detecting intrusions and anomalies. Rather, it provides operationally-focused technological and analytical support, giving experienced security analysts the tools they need to understand the traffic on their network. The highly efficient tools provided by the CENTAUR system have helped DoD analysts keep up with the rapidly increasing (1) traffic levels on the NIPRNet and (2) number of threats and attacks against DoD systems. In addition to the built-in analysis tools, the software suite has been designed to serve as an infrastructure on top of which people can add new capabilities and views with relative ease. A major example of this is the scan detection and analysis system currently being tested.

Point of Contact

Andrew Kompanek
CERT/NetSA Technical Lead, DoD Projects
ajk@cert.org
(412) 268-9744

Rex Brinker
SEI/CERT Project Manager, Information Assurance
rbrinker@sei.cmu.edu
(412) 268-7722

Discussion

Introduction to CENTAUR

The CENTAUR system provides users with flexible and efficient access to and analysis of a repository of detailed data regarding network traffic handled by the border and backbone routers on the NIPRNet. Data is gathered on all network traffic processed by the

- NIPRNet border (gateway) routers, representing the vast majority of traffic to/from the NIPRNet and the rest of the Internet, and
- NIPRNet backbone routers (called core and hub routers), representing the vast majority of traffic to/from NIPRNet enclaves.

CENTAUR provides near-real-time collection and storage of flow-level summaries of this traffic. In this context, flows are uni-directional aggregations of packets sharing the same transport layer protocol (e.g., TCP, UDP), source address and port, destination address and port, and arriving reasonably close together in time. Cisco's NetFlow is probably the best known system for reporting flows. Flow-level traffic data does not include payload, and is far less voluminous than packet-level data (even packet summaries without payload). The stored data files contain one record for each recorded flow, with the following data fields:

Report Documentation Page			Form Approved OMB No. 0704-0188	
<p>Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p>				
1. REPORT DATE JUN 2005	2. REPORT TYPE	3. DATES COVERED 00-00-2005 to 00-00-2005		
4. TITLE AND SUBTITLE The CENTAUR System: Helping to Protect the NIPRNet		5a. CONTRACT NUMBER		
		5b. GRANT NUMBER		
		5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S)		5d. PROJECT NUMBER		
		5e. TASK NUMBER		
		5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Carnege Mellon University, Software Engineering Institute, Pittsburgh, PA, 15213		8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)		
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited				
13. SUPPLEMENTARY NOTES				
14. ABSTRACT				
15. SUBJECT TERMS				
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 3
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified	19a. NAME OF RESPONSIBLE PERSON	

source IP address, source port, destination IP address, destination port, protocol, number of packets, number of bytes, logical OR of each TCP flag across packets, start time, duration, and sensor ID.

The CENTAUR system maintains this data on NIPRNet traffic dating back as far as April 2002. CENTAUR is currently in routine operational use by dozens of DoD information security analysts at JTF-GNO/NetDefense (formerly DoD-CERT), NSA, and other DoD components. Recent DoD adjustments to reporting chains (i.e., Operational Control (OPCON) for Service NOSCs to the JTF-GNO) make it especially prudent to make access available to the Services (e.g., AFNOSC), regional CERTs (e.g., CONUS), and others as appropriate. Recent usage trends include more than 1K analysis queries being run per day. Moreover, the historical record of NIPRNet traffic is actively used by the analysts: approximately 40% of queries access data more than one month old, including substantial numbers going back more than six months and an average of several queries per day going back even more than one year into the past.

Even using flow-level summaries of NIPRNet traffic, the volume of data gathered and stored is astounding. A typical weekday yields more than 4B flow records from the border routers, and a similar volume from the backbone routers; together these consume approximately 200 GB of disk storage per day. Although other systems for processing network flow data are available, they are not suitable for this volume of data. In order to support the requirements of DoD information security analysts for rapid responses to queries involving these vast amounts of data, along with requirements for affordable storage and processor costs, a customized system was clearly needed. In response, the CERT® team at the Software Engineering Institute¹ of Carnegie Mellon University, in close collaboration with DISA (and later JTF-GNO) and NSA, developed the SiLK² suite of software which forms the heart of the CENTAUR system. This software provides reliable data transmission and storage capabilities as well as a powerful set of analysis tools. The SiLK suite – and indeed all of CENTAUR (e.g., including the computing platforms) – were designed for rapid query response (data retrieval and processing), compactness in data transmission and storage, scalability, flexibility, adaptability, reliability, and to specifically support network security analysis.

CENTAUR / SiLK Analysis Capabilities

The SiLK suite provides CENTAUR with powerful and flexible capabilities to perform both current and retrospective analyses of NIPRNet traffic data. It is not yet another intrusion / anomaly detection system. Rather, we have provided operationally-focused technological and analytical support, giving experienced security analysts the tools they need to understand the traffic on their network. The SiLK analysis tools are general-purpose and do not assume or impose any particular model of security analysis. Moreover, the tools have been developed, tested, and refined in close collaboration with DoD network security analysts. The SiLK analysis tools provide Unix-like commands with the following types of functionality: selecting (a.k.a. filtering), displaying and sorting, summarizing in numerous ways, and manipulating aggregate data structures (namely, sets and bags). Multiple commands can be piped together for complex filtering and other processing, while only needing to read the packed binary data once from disk.

Numerous DoD network security analysts use the CENTAUR system on a daily basis to better understand traffic on the NIPRNet and to better defend it against attacks. Dozens of routine analyses are conducted on a daily or hourly basis. These are based upon scripted invocations of the SiLK analysis tools, set to run at predetermined times. These routine analyses

- identify malicious code activity to/from DoD systems (e.g., worm propagation attempts from inside or outside the NIPRNet)
- identify problems with network configurations and defenses (e.g., failures of ingress access control lists (ACLs) on the border routers)
- provide timely alerts suggesting emerging or evolving threats (e.g., lists of most active external sources; lists of external sources exclusively contacting pre-defined critical NIPRNet addresses; lists of most active ports)

¹ The Software Engineering Institute (SEI) is a federally funded research and development center sponsored by the U.S. Department of Defense and operated by Carnegie Mellon University.

² SiLK is an acronym for System for Internet Level Knowledge. In addition, the capitalized SLK is in memory of Suresh L. Konda, Ph.D., who was the visionary founder of the project.

- provide timely alerts for new traffic patterns, possibly indicating exploit attempts and malicious code (e.g., traffic volume by port compared to previous day, previous 30 day average, etc.; number of unique source and destination addresses by port)

The creation, centralization, and automation of these dozens of routine analysis runs have substantially enhanced the ability of information operations specialists to understand and defend the NIPRNet.

The SiLK tools are also used on a daily basis to support more ad-hoc analyses. Often these investigations are the result of a specific event, tip, or incident report. By focusing on pertinent data, these analyses lead to improved understanding of the situation and answer specific questions. In addition, they sometimes require accessing traffic data going back months into the past. For example, when a DoD system is found to have been compromised (e.g., root-level intrusion) an analyst will often identify all other NIPRNet addresses communicating with the compromised system because they may have also been attacked or may be the predecessor in the successful compromise. Examination of the flow records can help determine which systems may be affected. As another example, when DoD analysts learn from another source about a new vulnerability or exploit tool, they generally look for any evidence of its application to NIPRNet systems (often based on specific protocol and port combinations, flow sizes, etc.); they may then also set up ongoing analyses to watch for this potential activity.

CENTAUR Analysis Capabilities Built on Top of SiLK

In addition to the built in analysis tools, the SiLK suite has been designed to serve as an infrastructure on top of which people can add new capabilities and views with relative ease. A major example of this is the MISSILE³ scan detection and analysis system currently being tested. Scanning is frequently used by potential adversaries to identify potentially exploitable vulnerabilities of NIPRNet systems. In fact, the NIPRNet is subjected to hundreds of thousands of scan events every day. Network defenders need (1) a comprehensive, integrated view of scanning activity across the NIPRNet, as well as support to (2) help identify higher-risk scans, (3) identify hosts at greater risk of compromise, and (4) detect internal sources of scans. MISSILE incorporates a new approach to scan detection that is based on flows, is multi-dimensional and extensible, and provides the probability that selected traffic contains a scan. The MISSILE system includes scan detection, a scan database, and query tools to support the analysis of scans – all aimed at meeting the needs identified above. It is also expected to greatly reduce the burden of reporting scans, as is required by CJCSM 6510.01. Operational deployment of the system will automate this reporting task, improve the accuracy and completeness of reporting, and free up valuable analyst time so they can focus on actually defending the network.

In addition to the MISSILE system, other capabilities have been built on top of the SiLK infrastructure. For example, NSA analysts have developed prototype systems to identify KaZaA usage and to identify NetThief activity. Given these successful experiences, it is expected that many other valuable analysis tools will be developed for the CENTAUR system.

Conclusion

In conclusion, the CENTAUR system has been developed to help DoD information operations security analysts better understand and defend the NIPRNet. It has been deployed and used successfully (and increasingly) over the past 2+ years by dozens of DoD network security analysts. CENTAUR has helped them to keep up with the rapidly increasing traffic levels and number of threats and attacks against DoD systems. Interested DoD organizations are invited to contact Capt. Jeffrey Jaime (jeffrey.jaime@tic.dod.mil) at the JTF-GNO/ATU for further information about CENTAUR. In addition, the capabilities provided by CENTAUR are currently being extended to the SIPRNet. Finally, most of the SiLK software suite is available to the Internet community as open source software (<http://silktools.sourceforge.net>, silk-help@cert.org). Several of these non-DoD users are helping to test and improve the tools. Everyone – including the DoD – benefits from improvements in Internet security overall, as fewer systems will be available to attackers.

³ MISSILE is an acronym for Multiple Indicator Scoring for Scan Identification by Likelihood Estimation, and is being developed by the CERT Network Situational Awareness team at the Software Engineering Institute of Carnegie Mellon University.